Curation and Analysis of a Rodent Uterotrophic Database: Insights on Data Quality and Reproducibility

P Ceger[1], N Kleinstreuer[1], D Allen[1], X Chang[1], J Strickland[1], Q Zang[1], J Hamm[1], E Philips[1], B Jones[1], W Casey[2]

[1]ILS/NICEATM, RTP, NC, USA; [2]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

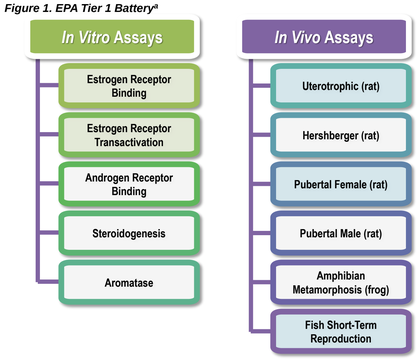
Abstract

High-quality *in vivo* reference data are critical to understanding the biological relevance of Tox21 and ToxCast *in vitro* assay data. The rodent uterotrophic bioassay, validated by OECD as a short-term screening test for assessing the estrogenic potential of chemicals, is included in the EPA's Endocrine Disruptor Screening Program as an *in vivo* Tier 1 screen. We performed a comprehensive literature review for uterotrophic bioassays conducted on 1812 chemicals in the EPA ToxCast screening program. Over 700 articles were identified as potentially relevant. Protocols used in each article were evaluated by two independent reviewers for conformity to six predefined criteria based on EPA and OECD uterotrophic test guidelines, with overall compliance determined by consensus. Studies meeting all criteria were considered guideline-like (GL). Information on 442 GL bioassays extracted from 92 articles and containing data for 98 ToxCast chemicals was compiled into a database of uterotrophic outcomes. The database includes data on 42 descriptors, including species/strain, number of animals per group, route of administration, duration of dosing, number of doses, maximum dose tested, lowest effect level, and test outcome. The immature rat model was used for 75% of the reported studies, with 73% of these using injection as the route of administration. Active outcomes were more common across rat models (74% active) compared to mouse models (36% active). Of the 70 chemicals in the database with at least two reported GL uterotrophic bioassays, 18 (26%) had discordant outcomes, many of which may be attributable to differences in study design (e.g. injection vs. oral dosing). This database provides a valuable resource for evaluating the performance of *in vitro* assays that measure key events in the estrogen receptor signaling pathway. *(Data in poster abstract have been updated to reflect the most recent analyses.)*

Introduction

- U.S. (7 U.S.C. 136, 110 Stat 1613) and international regulations require the testing of chemicals for the detection of potential endocrine activity. In the United States, this testing is conducted in the U.S. Environmental Protection Agency (EPA)'s Endocrine Disruptor Screening Program (EDSP).
- As many as 10,000 chemicals may lack sufficient testing data, with several hundred new chemicals being added each year (EPA 2011a, EPA 2012).
- The EDSP developed a two-tiered strategy for identification of endocrine-active chemicals. Tier 1 testing (**Figure 1**) includes *in vitro* and *in vivo* screens intended to identify chemicals requiring further testing to confirm endocrine activity.

Figure 1. EPA Tier 1 Battery[a]



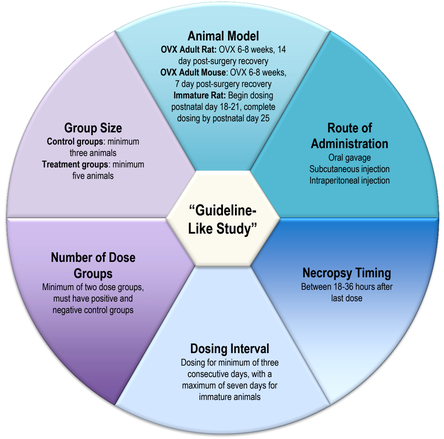
[a] Assays in shaded boxes are capable of detecting estrogenic activity.

- Using current methods, it could take decades to screen all 10,000 chemicals currently identified for Tier 1 screening (EPA 2014). The EPA ToxCast™ and the U.S. interagency Tox21 programs are working to address this problem. Both programs use quantitative high-throughput screening (qHTS) assays to assess how human biology is impacted by exposure to chemicals and identify exposures that are most likely to lead to adverse health effects.
- In order to use qHTS methods to prioritize and screen chemicals, it must be demonstrated that these methods are scientifically sound and can effectively detect endocrine bioactivity via receptor-mediated interactions and broader endocrine pathway effects.
- In support of this goal, the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) assembled a comprehensive database of high-quality *in vivo* reference data for endocrine-active chemicals.
  - This database focused on estrogenic effects because more data is available on chemical effects on estrogen signaling, from both *in vivo* and *in vitro* tests, than is available for other endocrine signaling pathways.
- The comprehensive database was created using the uterotrophic bioassay as the reference test method because:
  - It is a short-term assay that uses changes in uterine weight in response to chemical exposure, based on the uteruses' natural growth response to estrogens.
  - Uterine weight assays are represented in the scientific literature from the 1930s (Korenchevsky et al. 1935) to present (Yao et al. 2014).
  - It is one of the most extensively utilized estrogen detection assays and has been used to test thousands of chemicals (National Cancer Institute 1968).
  - It is used in the EDSP Tier 1 as a short-term *in vivo* bioassay to evaluate the ability of a chemical to elicit a biological response similar to that of natural estrogens (EPA 2011b; OECD 2004).

NICEATM Process Overview

- A set of 1812 chemicals was identified as having been tested using 18 *in vitro* Tox21 and ToxCast estrogen receptor (ER) assays.
- NICEATM conducted comprehensive literature searches to identify published studies in which the 1812 chemicals were tested in the uterotrophic assay.
  - Searches used chemical name, chemical synonyms, and Chemical Abstracts Service Registry Number as the search terms.
  - Modifier terms including "uterotrophic assay", "uterotrophic", "uterotropic" (a common alternative) and "uterine weight" were used to identify articles of interest.
- More than 1000 articles were identified by the initial searches; of these over 700 were deemed potentially relevant.
- Potentially relevant articles were downloaded and detailed study information was extracted for 42 descriptors, including:
  - Species and strain of test animal
  - Dose administration route
  - Age at first dose
  - Ovariectomy (OVX) status and age at which OVX was performed
  - Dosing length/frequency, number of doses, time of necropsy
  - Number of animals per dose group
  - Positive/negative controls
  - Maximum dose tested
  - Reported bioactivity for the dose range tested: "active" with a corresponding lowest effective level or "inactive" with a corresponding highest dose tested
- Each study was reviewed; data were extracted and evaluated for how well they met study design requirements described in accepted test guidelines for the uterotrophic assay (EPA OCSPP 890.1600 [EPA 2011b] and Test Guideline 440 issued by the Organisation for Economic Co-operation and Development [OECD 2004]). These requirements, termed minimum study criteria, are shown in **Figure 2**.

Figure 2. Minimum Study Design Criteria



- For each study evaluated, a score of 0 (no) or 1 (yes) was recorded for each of the six minimum study criteria based on whether the study protocol fulfilled that particular requirement. These scores were added to yield a total score for each study, and only those studies with a score of 6 were considered guideline-like (GL).
  - Protocols used in each article were evaluated by two independent reviewers for minimum study criteria, with overall compliance determined by consensus.
- The resulting database of uterotrophic outcomes currently contains data for 98 ToxCast chemicals, with information from 442 GL studies extracted from 92 publications.
  - Data from estradiol and ethinyl estradiol, chemicals that are commonly used as positive controls, were excluded from analysis. This was because of the large number of results for these chemicals and also because of the inherent bias associated with their inclusion (negative results would be treated as a failed positive control, likely resulting in the entire study not being reported).
  - Removal of data for estradiol and ethinyl estradiol left 357 GL studies, covering 96 chemicals.

Distribution of Outcomes Across GL Uterotrophic Study Designs

- The GL studies conformed to six basic study designs (**Table 1**) that used rats or mice, administered test chemical orally or by injection, and used OVX or immature animals.
- The majority (75% [268/357]) of studies were conducted using the immature rat model, followed by the OVX rat (12% [42/357]) and OVX mouse (13% [47/357]) models.
- Injection was the more common (71% [254/357]) route of administration across all experimental models. For 99% (251/254) of these studies, test chemical was administered by subcutaneous injection; a few studies used intraperitoneal injection.
  - 73% (196/268) of immature rat studies used injection as the route of administration.
- A breakdown of results by study design is presented in **Table 1**. Active outcomes were more common in rat studies (73% [227/310]) compared to mouse studies (36% [17/47]).

Table 1. Distribution of Uterotrophic Outcomes by Study Design[a]

	Im_Rat Inj	Im_Rat Oral	OVX_Rat Inj	OVX_Rat Oral	OVX_Mouse Inj	OVX_Mouse Oral
Percent of total studies using this protocol	0.55	0.20	0.09	0.03	0.07	0.06
Active[b] (number)	142	52	29	5	11	6
Inactive (number)	55	19	3	5	14	16
Active (percent)	0.72	0.73	0.91	0.50	0.44	0.27
Inactive (percent)	0.28	0.27	0.09	0.50	0.56	0.73

Abbreviations: Active = the number or percent of experiments using this design that reported substances as active; Im = immature; Inactive = the number of experiments reporting substances as inactive; Inj = injection; OVX = ovariectomized.

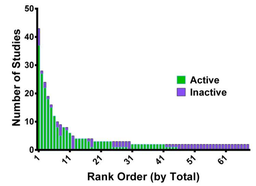
[a] Table includes data for 357 guideline-like uterotrophic studies, covering 96 chemicals. Data from chemicals commonly used as positive controls (such as estradiols) were excluded. This was because of the large number of results for these chemicals and also because of the inherent bias associated with their inclusion (negative results would be treated as a failed positive control, likely resulting in the entire study not being reported).

[b] The range of chemicals tested in these studies is neither randomly nor uniformly distributed with respect to uterotrophic bioactivity. Therefore, the performance of a particular study design, especially one with a small number of studies (e.g., OVX_rat_injection), could be heavily influenced by a single publication from one laboratory testing multiple chemicals using that particular study design.

Reproducibility of Uterotrophic Outcomes

- Seventy chemicals (excluding positive controls) in 357 studies had at least two reported uterotrophic assay results.
- The distribution of active and inactive results across chemicals is presented in **Figure 3**.

Figure 3. Rank Order Distribution of Uterotrophic Outcomes



- Of the 70 chemicals tested in multiple GL studies, 18 (26%) had at least one study with a discordant result.
  - Details of discordant results are listed in **Table 2**.
  - Discordance may be due to a number of factors, including differences in overall study design or the range of doses tested in each study (for example, when a substance that is only positive at high concentrations is not tested within its active range).
  - Ten substances (underlined and shaded in grey in **Table 2**) had discordant outcomes that may be attributable to differences in study design.
- Butylparaben provides an example of how uterotrophic outcomes could potentially be impacted by study design (**Figure 4**).
  - All active results were reported in the three study designs using s.c. injection as the route of administration (immature rat, OVX rat, OVX mouse).
  - Conversely, inactive results were reported for both study designs that used oral dosing (immature rat, OVX mouse).

Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contract HHSN27320140003C.

The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

A summary of NICEATM activities at SOT 2015 is available on the National Toxicology Program website at http://ntp.niehs.nih.gov/go/742110.

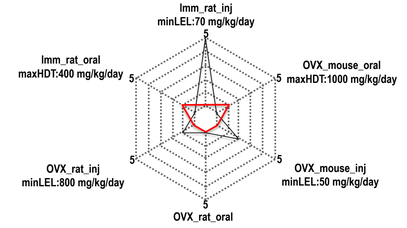
Table 2. Chemicals with Discordant Uterotrophic Results in GL Studies

CASRN[a]	Chemical Name	Total Studies	Number of Active Results	Number of Inactive Results
80-05-7	Bisphenol A	43	37	6
446-72-0	Genistein	28	27	1
72-43-5	Methoxychlor	19	18	1
789-02-6	o,p'-DDT	16	15	1
94-26-8	Butylparaben	10	8	2
56-53-1	Diethylstilbestrol	9	8	1
104-40-5	4-n-Nonylphenol (linear, para)	9	5	4
140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	4	3	1
120-47-8	Ethylparaben	4	1	3
119-61-9	Benzophenone	3	1	2
99-76-3	Methylparaben	3	1	2
56-55-3	Benz(a)anthracene	3	1	2
1806-26-4	4-Octylphenol	3	1	2
94-13-3	Propylparaben	3	1	2
52645-53-1	Permethrin	2	1	1
50-55-5	Reserpine	2	1	1
520-36-5	Apigenin	2	1	1
486-66-8	Daidzein	2	1	1

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; GL = guideline-like.

[a] Shaded and underlined chemicals had discordant uterotrophic outcomes in study designs that differed significantly from one another. The remaining chemicals had discordant results reported in assays with similar study designs.

Figure 4. Concurrence of Study Design with Uterotrophic Outcomes for Butylparaben[a]



Abbreviations: maxHDT = maximum high dose tested; minLEL = minimum lowest effect level.

[a] Studies with active results are connected with black lines; studies with inactive results are connected with red lines.

References

EPA. 2011a. The Incorporation of *In Silico* Models and *In Vitro* High Throughput Assays in the Endocrine Disruptor Screening Program (EDSP) for Prioritization and Screening. Summary Overview. A Part of the EDSP Comprehensive Management Plan [Internet]. Washington, DC:U.S. Environmental Protection Agency. Available: http://epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf.

EPA. 2011b. Standard Evaluation Procedure Uterotrophic Assay OCSPP 890.1600 [Internet]. Washington, DC:U.S. Environmental Protection Agency. Available: www.epa.gov/endo/pubs/toresources/seps/Final_890.1600_Uterotrophic_Assay_SEP%209.22.11.pdf.

EPA. 2012. Endocrine Disruptor Screening Program Universe of Chemicals and General Validation Principles [Internet]. Washington, DC: U.S. Environmental Protection Agency. Available: http://www.epa.gov/endo/pubs/edsp_chemical_universe_and_general_validations_white_paper_11_1 1.pdf

EPA 2014. Integrated Bioactivity and Exposure Ranking: A Computational Approach for Prioritization and Screening Chemicals in the Endocrine Disruptor Screening Program [Internet]. Washington, DC: U.S. Environmental Protection Agency. Available: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2014-0614-0003

Food Quality Protection Act of 1996. 7 U.S.C. 136. Public Law 104-170.

Korenchevsky V, Dennison M, Simpson SL. 1935. The prolonged treatment of male and female rats with androsterone and its derivatives, alone or together with oestrone. Biochem J. 29(11): 2534–2552.

OECD. 2004. Test No. 440. Uterotrophic Bioassay in Rodents: A Short-Term Screening Test for Oestrogenic Properties. OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Parism France: OECD Publishing. Available: http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents_9789264067417-en

National Cancer Institute. 1968. Endocrine Bioassay Data. Bethesda, MD:National Institutes of Health.

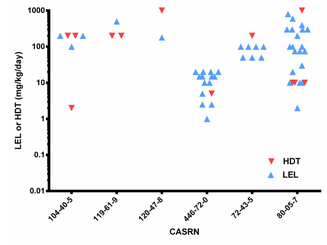
Safe Drinking Water Act Amendments of 1996. 110 Stat 1613. Public Law 104-182.

Yao PL, Ehresman DJ, Rae JM, Chang SC, Frame SR, Butenhoff JL, et al. 2014. Comparative *in vivo* and *in vitro* analysis of possible estrogenic effects of perfluorooctanoic acid. Toxicology. 326: 62–73.

Potential Effect of Dose Range in Discordant Studies

- To evaluate whether the highest dose tested (HDT) may not have been high enough to detect a uterotrophic effect, the HDT for inactive outcomes was compared to the lowest effect level (LEL) reported for active outcomes (**Figure 5**).
- To simplify the analysis, the only studies examined were those using the immature rat injection design, which produced the most discordant results.

Figure 5. LEL/HDT Comparison for Chemicals with Discordant Outcomes in the Immature_Rat_Injection Study Design[a]



Abbreviations: CASRN = Chemical Abstracts Service Registry Number; HDT = highest dose tested; LEL = lowest effect level.

[a] Of the eight substances in non-shaded rows in **Table 2**, *o,p'*-DDT had no discordance among studies using the immature rat injection design. Diethylstilbestrol had discordant results among multiple immature rat injection studies but the doses used could not all be converted to mg/kg/day.

- For discordant results among studies using the same study design, it was common for the HDT yielding inactive results to be greater than the LEL that produced an active result. For example, bisphenol A (80-05-7) was active by s.c. injection at 2 mg/kg/day in one immature rat injection study, but inactive at 1000 mg/kg/day in another study using the same model.
- Thus, it does not appear that the highest dose tested is a cause of discordant outcomes.
- Further investigation into study design, e.g., strain differences, etc., may provide additional information.

Conclusions

- High-quality *in vivo* reference data are critical to understanding the biological relevance of Tox21 and ToxCast qHTS data.
- NICEATM conducted a large literature review to compile a database of high-quality *in vivo* uterotrophic assay reference data from GL studies.
- Preliminary evaluation of GL studies found that:
  - The majority (75%) of studies were conducted in immature rats, with the most common route of exposure being injection.
  - Active outcomes were more prevalent in studies using rat models, with 50–91% of chemicals being active in the OVX rat model, 72–75% of chemicals being active in the immature rat model, and 27–44% of chemicals being active in the OVX mouse model.
  - Chemicals were less likely to be active by the oral route in the OVX rat and OVX mouse models, but not in the immature rat model.
- For 70 chemicals tested in more than one GL uterotrophic study, 18 had discordant results.
  - For 10 of the chemicals, the discordance may be attributable to differences in study design.
  - Eight chemicals had discordant results across multiple studies using the same basic protocol (e.g., immature rat injection).
  - An initial evaluation of data from immature rat injection studies for the eight chemicals with discordant results in these studies indicates that the discordance is not due to having an insufficiently high dose range to observe chemical effects.
  - Further evaluation is ongoing and may yield additional insights into the reasons for discordant results.
- This database provides a valuable resource for evaluating the performance of *in vitro* assays that measure key events in the estrogen receptor signaling pathway.
- The uterotrophic database has been used for an evaluation of ToxCast data. The complete evaluation with supporting material are available (EPA 2014).